DOCUMENT RESUME

ED 435 761 TM 030 413

AUTHOR Howell, Lorrie A.; Brocato, D. Kay; Patterson, Kim; Bridges, Tricia

TITLE An Evaluation of Performance Task Instruction in the Classroom.

PUB DATE 1999-11-00

NOTE 25p.

PUB TYPE Reports - Research (143)

EDRS PRICE MF01/PC01 Plus Postage.

DESCRIPTORS *Academic Achievement; Educational Improvement; *Elementary School Teachers; Elementary Secondary Education; *Performance Based Assessment; Performance Factors; *Professional Development; School Districts; *Teaching Methods

IDENTIFIERS Mississippi

ABSTRACT

This study attempted to provide insight into the condition of one Mississippi school district that has engaged in a comprehensive effort toward instructional improvement centering around performance task instruction by examining how teaching with performance units affects several aspects of the instructional program. In the district studied, the Program of Research and Evaluation for Public Schools (PREPS) Integrated Assessment module of staff development, developed by curriculum specialists working with the Mississippi State Department of Education and a representative of the Riverside Publishing Company, was used to train teachers in performance unit construction and teaching strategies. Student achievement data show that when mathematics performance assessment scores and language performance scores are analyzed separately, the hypothesis that performance task instruction would increase scores was not supported. When total performance scores are analyzed, there is a significant improvement in scores for 1999 over those of 1998. Data do not show a relationship between performance assessment scores and core battery scores. Teachers indicated that performance-based instruction and assessment had improved their own performance and student performance overall. The school district has continued its commitment to performance-based instruction and assessment. (Contains 17 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

Running head: PERFORMANCE TASK INSTRUCTION

ED 435 761

An Evaluation of Performance Task Instruction in the Classroom

Lorrie A. Howell, D. Kay Brocato, Kim Patterson, and Tricia Bridges

Mississippi State University

BEST COPY AVAILABLE

TM030413

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

L. Howell

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

An Evaluation of Performance Task Instruction in the Classroom
Outline

I. Introduction

- A. Performance Assessment at National, State, District, and School Levels
- B. What is Performance Assessment
- C. Implementing Performance Based Instruction
- D. Supports and Limits of Performance Assessments
- E. Significance of the Study
- F. The Performance Assessment Professional Development Problem
- G. Answering Performance Instruction Questions

II. Method

- A. Subjects
- B. Procedure
- C. Instruments
  - 1. Post Training Questionnaire
  - 2. ITBS
  - 3. TAP
  - 4. Performance Assessments
- D. Experimental Design
- E. Data Analysis

III. Results

- A. Quantitative Results
- B. Questionnaire Results: Themes in Qualitative Data
  - 1. Confidence in Test Taking.
  - 2. Strategies Used During the Test.
  - 3. Changes in Teachers' Practices.
  - 4. Developing Higher Order Thinking Skills.

IV. Discussion

V. Conclusion

VI. References

VII. Appendix: Summary of Teacher Responses from the Post-treatment Questionnaire

An Evaluation of Performance Task Instruction in the Classroom

Introduction

Current interest in performance-based assessment has lead the future of student assessment to a critical point where decisions must be made. Many states are now rethinking their current assessment systems. Developing, and eventually implementing, an assessment system that is non-multiple choice is difficult, time-consuming, and extremely controversial. However, education reformers support the belief that assessment should be generative and render an active production of the student's work. Testing should be meaningful instead of testing what is easy to test and easy to score. According to Khattri, Reeve, and Kane (1998), the theory at the cornerstone of assessment reform states that "performance assessments must require students to structure the assessment task, apply information, and construct responses, and in many cases, students must also be able to explain the processes by which they arrive at the answers" (p. 2).

National, State, District, and School Levels

National agencies, projects, and organizations, [e.g., New Standards Project (NSP), the Coalition of Essential Schools (CES), and the College Board's Pacesetter program] have become involved in the current educational reform. These organizations and projects are pushing for assessment, curricular, and instructional reform. Performance assessment is a focal point of all of these projects (Hakel, 1998). National administrators realize the importance of this reform, and made their commitment evident by passing Goals 2000: Educate America (P. L. 103-227). The agenda of Goals 2000 is to have education completely reformed by the year 2000 and for teachers to have access to workshops and programs to further their professional development. Goals 2000 gave, and still is giving, states the ability to develop a standards-based education by granting money to fund the changes. Title I, another national program is pushing for performance-based assessment to be an optional measure beyond the norm-referenced, multiple-choice testing (Borko, Mayfielf, Marion, Flexer, & Cumbo, 1997; Hakel, 1998).

Individual states and districts have also gotten involved in the performance-based assessment reform movement. As of 1994, 38 states were using writing samples to measure

writing proficiency, and 25 used performance-based items in their assessment systems. States such as Vermont, Maryland, and Kentucky have been among the first states to fully implement performance-based assessment systems in writing and mathematics. However, these reforms have not been implemented in these states without receiving criticisms from parents who feel the assessments contain distasteful material and are intrusive (Khattri, Reeve, & Kane, 1998).

Districts and individual schools have also begun to implement performance-based assessments. Although the changes have not been as noticeable, sometime reform at this level is easier because it is on a smaller scale. Some schools have implemented graduation examinations based on performance assessments. However, these changes in assessment require more than simply formulating a test. Curriculum and instruction must also by revised in order to align with current testing systems. Teaching methods must become more holistic in nature. Many current preservice teaching programs do not train teachers in this. Therefore teachers must look elsewhere for training (Khattri, Reeve, & Kane, 1998).

Mississippi's use of Performance Assessments ushered in instructional changes considered to be pedagogically sound. Hinckley (1999), president of MENC, explains why he feels performance assessment is a worthy endeavor. Hinckley claims grading performance allows less subjectivity and is more fair, accurate, and consistent. Students can show what they know and can do so more clearly to evaluators. Performance assessment tells about the success of teaching and makes students more accountable. For these reasons, performance-based assessments have received renewed interest over the past 10-15 years (Hakel, 1998). The recent turn from standardized testing toward a performance task approach is widely viewed as a way to improve the reputation of schooling in America. Khattri, Reeve, and Kane (1998) claim that the focus on performance assessment exists for three reasons. First, front line educators, especially in the k-12 arena, stand united against accountability based solely on multiple choice tests. Second, the rapid increase of information and mass access to that information has led educators to more readily support the constructivist theory of learning. Constructivist learning supports the idea that analysis, synthesis, and evaluation are highly intricate processes which require students

to have a personal learning purpose, to use prior knowledge, and to engage in specific content related problems or tasks (Henderson 1996). Brooks and Brooks (1993) explain constructivist theory to be grounded on the following questions:

> Can students demonstrate comprehension of concepts, not memorized information?
>
> Can students solve problems, instead of simply rotely following procedures?
>
> Can students inquire into complex issues, instead of simply parroting rehashed beliefs?

Khattri, Reeve, and Kane (1998) report a final reason for the focus on performance assessment. They state that the working world has become increasingly concerned about American schools producing a work force who is highly productive and globally competitive. Large scale standardized performance assessment is believed to be some part of the solution for transforming schools into places that produce people who can produce.

What is Performance Assessment

Performance assessment items are multi-stepped tasks which require students to structure their own learning activity. To create this structure, students are required to apply information, knowledge, and previous learning. All performance tasks require student-constructed responses. Analysis of the task and a self study of a learner's role in the tasks are often a required part of performance assessment. This means that students are required to explain the process by which they arrived at the answer. Performance tasks are often evaluated using rubrics which set a hierarchy of performance standards and expectations. Rubrics are said to engage and empower students in the learning process. Khattri, Reeve, and Kane (1998) describe several types of performance tasks: in-demand tasks, which are quick, time-limited student constructed response items; extended tasks, which are similar to in-demand tasks but allow time to think, research, and demonstrate mastery; demonstrations or presentations; portfolios, which are a container of various performance task products chosen by the student that demonstrate development or growth of self; and observations by teachers or experts of students in a natural setting

demonstrating progress. These observations are often used by preschool teachers to assess student progress. For example, early-learners show impulse control by working and playing well with others. Students simply perform this task of working and playing well with others which shows mastery of this objective while teachers document said mastery. Hence, performance is assessed as the task is performed in real situations.

Implementing Performance Based Instruction

Madaus and O'Dwyer (1999) supply three ways in which performance-based assessments can be implemented in the classroom. Teachers can ask the student to give an oral or written answer to a series of questions. Short answers and essays are examples of this. Teachers might also require the student to produce a product such as a portfolio or a research paper. A third way in which performance-based assessment can be implemented by the teacher is to request the student to perform an act to be evaluated using set criteria. Examples of this include reading a book aloud or conducting a chemistry experiment. Performance-based assessment is already in use in many classrooms in this manner; however, until recent years these tasks were not thought of as a valid method of assessment.

Bringing performance-based assessment into the classrooms means changing teaching methods. Assessment methods will go far beyond the typical end-of-the-chapter tests. Teachers who use performance-based assessment will require the students to demonstrate what they have learned. As students prove their competence, teacher expectations will rise. It is believed that bringing performance-based assessment into the classroom will bring a host of instructional changes. The line between teaching and testing will be less clear. Subsequently, students will be more motivated to learn (Haertel, 1999).

Supports and Limits of Performance Assessments

Heartel (1999) gives some warnings about performance assessment in a *Phi Delta Kappan* report. He states that there is no simple solution to the issues of teaching, learning, and assessment and warns about placing too much dependence on performance. Heartel feels that even with assessments like the Iowa Tests of Basic Skills (ITBS) and Tests of Achievement and

Proficiency (TAP), using one measure of achievement for accountability is not sufficient. He goes on to claim that performance assessment does not effect sweeping educational reform. Still, he feels that performance should be used for daily classroom instruction because performance tasks are not timed in class, teachers can help with the tasks, and students are able to get more immediate feedback. Other skeptics claim that lack of time and poorly defined standards prohibit performance assessment from being widely useful. Also, even though evaluations like the ITBS/TAP are more likely to address various forms of intelligence, students with strengths in certain areas are likely to do better than students with intelligence strengths in other areas (Gardener, 1994).

Although performance-based assessment shows much promise, this type of testing also has limits. Many educators question the psychometric characteristic of performance assessment. Professionals must develop scoring rubrics that are objective and reliable. Sometimes the scoring of performance-based assessment poses the biggest obstacle in performance assessment. Educators also do not favor performance assessment because of its inability to be used for individual placement decisions for students (Elliot & Fuchs, 1997). In an unreferenced article, Givens (1997) states that performance-based assessment lacks adequate research to be effectively implemented in the classroom. Givens raises questions concerning scorers qualifications; scorers objectivity-- or lack there of; and scorers ability to determine the quality of response.

Another criticism comes from the teachers concerning time for preparation. Performance-based assessment requires more time to prepare the students for the assessment. A survey conducted with teachers in Kentucky and Maryland indicated students' increases in achievement from one year to the next was not due to the actual implementation of performance-based assessment in the classroom. It was due to the teachers spending more time in drill and practice and their awareness of what was expected of the students. Teachers stress more abstract skills while neglecting art, music, physical education, and mechanics (e.g., grammar and spelling). Teachers also criticized performance-based assessment because of the delay between test-taking

and receiving the scores. Because of this delay, teachers are unable to know how to prepare students for the next test. Although opponents believe that performance-based assessment adds value to education and instruction, they do not accurately measure students' knowledge or aptitude. Opponents believe that regardless of the amount of research conducted on performance-based assessment this will never be achieved (Givens, 1997).

One criticism of standardized, multiple-choice tests is cultural bias. It is believed that multiple-choice tests favor students who normally score higher on tests. However, education reformers believe that performance-based assessment will yield scores that are more equal between students from rich communities and students from poor communities. If students from poorer communities are given a chance to actually demonstrate their knowledge, they will perform just as well as students from more wealthy communities (Haertel, 1999).

On the contrary, Miller (1999) reports that instructional improvement is possible through performance assessment, and that staff development can help teachers understand how to use a performance focus for improving student learning. He reports that Performance brings writing to other content areas and helps students think logically and apply existing knowledge to new situations.

Performance-based assessment is appealing to many educators for obvious reasons. These educators see the many changes across the nation taking place in schools that have implemented performance-based assessments. Educators have also found moderate correlations between performance-based assessment scores and multiple-choice test scores. Implementing performance assessment in the classroom could possibly have a carry-over effect and also improve students' scores on standardized, multiple-choice tests. Performance-based assessment is also favored because it has high content validity. Multiple-choice tests' content validity has always been a point of controversy. Many times, especially with national test, it is difficult to make sure multiple-choice tests are tailored for the curriculum. Another positive aspect of performance assessment is that although it may contain more difficult tasks, educators have

found that students actually favor performance assessment over multiple-choice testing. Students become more excited and motivated to learn (Elliot & Fuchs, 1997).

Significance of Study

During the 1994-1995 school year, Mississippi administered the first round of Performance Assessments to all fourth through eighth grade students. State mandated assessment in the past had been limited to multiple-choice questions that required students to select answers from a list of given responses. Among other thinking tasks, Performance Assessments require students to write, to construct responses, and to explain answers. Teachers reported that students were overwhelmed with performance assessment items since familiar standardized test items had only required students to choose and bubble in answers. Educators could clearly see that skill and drill, worksheet mania would not prepare students to perform well on Performance Assessments. This shift in assessment type required changes to instruction in Mississippi classrooms. Teachers had to provide performance-based assignments, writing activities, and often group work for students. Teachers had to learn to score with rubrics and integrate technology in useful ways that would enhance performance. A large scale effort to change instructional strategies was necessary to realize two of the guiding principles of the Mississippi Board of Education, "A commitment to higher expectations for students, parents, teachers, and administrators at all levels./ A belief that reading is the most essential basic skill for lifetime success (State Report Card, 1995).

Later versions of the Mississippi State Report Card indicate that the Mississippi Department of Education holds the vision of a "Quality education for every child . . . every child a reader(1998, p. i)" and establishes focus areas of "reading, student advancement, teachers and teaching(1999, p. i)." Statewide usage of Performance Assessments— as opposed to only multiple choice tests— has made strong and consistent moves toward these goals. Hakel (1998) claims that being able to recognize a correct answer out of a list of choices is hardly equivalent to constructing a correct response. His criticism of a multiple choice test lies in the psychometric properties— particularly areas such as content validity. The problem lies not in predictions

based on the scores, but in description and diagnosis. Hakel says that selected response items may leave educators wondering about what students can do after instruction that they could not do before. Multiple choice tests never address issues concerning what the child did not learn. Simple multiple-choice tests are not likely to answer these questions (Hakel, 1998). Because of these questions and the realization that traditional tests might not answer these questions, education professionals are calling for reform. Portfolios and "authentic" performance-based assessments are receiving more attention (Hakel, 1998).

The Performance Assessment Professional Development Problem

One factor crucial to successful performance-based assessment is the presence of effective performance instruction professional development. The way in which curriculum and instruction currently stands, teachers train their students to perform on the norm-referenced, multiple-choice tests. With the new push for performance-based assessments, teachers must change their expectations of the students and their methods of teaching. Students are no longer only expected to circle the correct answer or fill in the correct bubble; they must now be able to construct their own responses, think critically, and have more than just basic factual knowledge (Khattri, Reeve, & Kane, 1998).

Teachers must develop professionally in order to learn how to design assessments, implement these new assessments and finally score the finished product. Most teachers involved in this reform attend assessment task and scoring rubric workshops (Martin & Miller, 1996). Muscella, Paget, and Gibbel (1994) suggests that performance assessment should be included as part of preservice training for future teachers. By doing so new teachers will strive to continually improve their teaching method through meaningful assessments.

Answering Performance Instruction Questions

This study attempted to provide insight into the condition of one Mississippi school district which has engaged in a comprehensive effort toward instructional improvement centering around performance task instruction. This research seeks to examine how teaching with performance units effects several aspects of one school districts instructional program. First, the

results of students' standardized core battery/multiple choice assessment scores were examined. Second, the results of students' standardized performance assessment scores were examined. Third, students' experience of test anxiety and feeling of success during the testing situation was examined from teachers' perspective. Finally, teachers' perception of using performance task instruction was studied.

Method

Subjects

Numerous Mississippi school districts contracted with PREPS for Integrated Assessment training. A north-central school district was chosen as the sample population because this district was considered the training site with the strongest treatment integrity for several reasons. First, only this district provided the training to all district teachers in grades 2-9. Second, instructional leaders expressed a need for help in teaching strategies since the school community had clearly related dissatisfaction with the districts' decline in performance during the last testing cycle. Third, district officials related a strong commitment to the instructional strategies espoused by PREPS Integrated Assessment in order to improve test scores. Finally, the staff development leaders continued to provide technical support to district teachers assisting with performance unit construction and performance teaching strategies. Sample size included 6 scores for mathematics and language arts performance assessment scores for each year. Total performance assessment scores involved a sample size of 12 for each year. Eighteen core battery scores were used for the correlation procedure. All nineteen teachers receiving the post training questionnaire responded.

Procedure

The treatment in this study is defined as the PREPS Integrated Assessment module of staff development. PREPS Integrated Assessment curriculum pieces were written by curriculum specialists who worked closely with the Director of Assessment from the Mississippi Department of Education and with Riverside Publishing Company's Mississippi assessment consultant. Riverside held the state contract for all state assessments given in grades 4-8 from 1994-1999. PREPS' Integrated Assessment curriculum specialist wrote instructional materials that were

BEST COPY AVAILABLE

directly aligned with assessment objectives and other standards as defined by content area professional powers (NAEP, NCTM, NCTE, ASCD, etc.). The training module is a four-day staff development experience for teachers of language arts and math in the 2nd through 9th grades. The training is designed to introduce teachers to instruction that is performance based. Performance units are defined through an exploration of the following components:

- multi-stepped tasks
- use of manipulatives
- presence of literature/text
- students' practice of returning to text
- student constructed responses/student writing
- checklists for students' self evaluation
- rubrics for external evaluation
- checklists' correlation to rubric

The introduction to teaching with performance tasks is followed with a step-by-step modeling of teaching using an original performance unit. Staff development leaders guide district teachers through one language arts and one mathematics performance unit exactly as the unit is designed to be worked through with students. Teachers are encouraged to be creative with the given units but are asked to keep units consistent with the definition of performance instruction. The final step of the PREPS Integrated Assessment model asks teachers to write an original performance unit. Each teacher receives two original performance units to use directly after the training module upon returning to the classroom.

Instruments

Post Training Questionnaire. The post training questionnaire asks participants to identify the location of the session they attended and the grade they teach. The participants are not asked

to give their names. The first six items on the questionnaire are a 1 through 5 Likert scale with five being outstanding and/or very valuable and 1 being poor and /or irrelevant. Participants are asked to rate the session according to quality of presentations, content and delivery; discussion and interaction with speakers; small group work; overall effectiveness; impact on developing performance units for classrooms; and impact on developing rubrics in classrooms. Participants are then asked to provide comments related to the Likert scale, especially the impact of the sessions on developing performance units and rubrics for their classrooms. The next two items on the questionnaire ask participants to describe the most effective part of the staff development series and make suggestions in the event that the sessions are presented again. Participants are then given a list of strategies presented during the sessions and are asked to indicate which strategies they tried in the classroom, then choose one of those strategies and describe how they implemented it and how it impacted their classroom. The last item on the questionnaire asks participants to think about the days when their students took the performance portion of the ITBS/TAP and describe things they noticed about their students as they completed the performance portion of the test. For the purposes of this study, only the portions of the questionnaire that were analyzed were the comments on training, implementation in the classroom, and teachers' perceptions of students taking the test.

ITBS. The ITBS is a norm-referenced test that is administered in grades 4 – 8 and consists of traditional multiple-choice items covering the content areas of reading, mathematics, and language arts. These standardized test measure student achievement in relation to the performance of a national sample. The national mean is a Normal Curve Equivolent (NCE) of 50, which corresponds to the 50th percentile.

TAP. The TAP is a norm-referenced test that is administered in grade 9. The TAP consists of traditional multiple-choice items that cover the content areas of reading, mathematics, and language arts. These standardized tests are used to measure student achievement in relation to the performance of a national sample. The national mean is an NCE of 50, which corresponds to the 50th percentile.

Performance Assessments. The performance assessment portion of the test consists of items requiring the students to construct responses or write their answers. Performance assessments require students to explain their answers making it possible to assess their thinking and problem-solving skills.

Experimental Design

This study follows a one group pretest-posttest experimental design. The dependent variable was the mean language arts and math performance assessment scores for 4th through 9th grade students from the 1996-1997(pretest) and then the 1997-1998(posttest) testing cycle. The most profound limiting factor of this study was the small sample size. Future studies should consider individual student scores for data analysis. Also the one-group pretest-posttest design has inherent internal validity threats of history, maturation, testing, instrumentation, and interaction of selection with other factors. External threats include interaction between testing and treatment and interaction between selection and treatment. The independent variable or experimental treatment is the PREPS Integrated Assessment staff development training module delivered to teachers prior to the 1997-1998 state testing cycle.

Data Analysis

The mathematics and language performance assessment scores were analyzed by using a one-way analysis of variance (ANOVA) procedure. The sample for these analyses included 6

scores for both of the years. Total performance assessment scores were also analyzed with a one-way ANOVA, which involved a sample size of 12 for each year. A Pearson product-moment correlation was also calculated to examine any relationship between core battery scores and performance scores for 1998 and 1999.

Quantitative Results

Table 1 lists the means and standard deviations of ITBS/TAP scores by test and year. Data for the Mississippi Mathematics Performance Assessment were analyzed using a one-way analysis of variance (ANOVA) procedure. The ANOVA indicated no significant differences for mathematics 1998 and 1999 scores, F (1, 10) = 1.680, $p > .10$. An ANOVA procedure was also used to analyze data for the language performance assessment and indicated that there were no significant differences between language scores for the years 1998 and 1999, F (1,10) = 2.751, $p > .10$. A one-way ANOVA was conducted to analyze any differences for total performance assessment scores. Results indicated a significant difference between total performance assessment scores, F (1, 22) = 3.741, $p < .10$.

Table 1

Means and Standard Deviations for Performance Assessment Scores

	1998		1999	
	Mean	**S. D.**	**Mean**	**S. D.**
Language Performance Assessment (N = 6)	**425.400**	**19.947**	**444.600**	**20.152**
Mathematics Performance Assessment (N = 6)	**406.083**	**26.987**	**423.950**	**20.297**
Total Performance Assessment (N = 12)	**415.742**	**24.772**	**434.275**	**22.094**

Mathematics, Reading, and Language Core Battery data were also analyzed to assess the presence or absence of a relationship between performance assessment scores and core battery scores. Table 2 shows increase in means and standard deviations for core battery scores. It was hypothesized that as performance assessment scores increased, core battery scores would increase and vice versa. A Pearson product-moment correlation was calculated to assess any relationship between these scores. For 1998 and 1999, there appears to be no statistically significant relationship between performance assessment scores and core battery scores ($r = 0.009$, $p > .10$; $r = 0.266$, $p > .10$; $r = 0.370$, $p > .10$; respectively).

Table 2

Means and Standard Deviations for Core Battery Scores

	1998		1999	
	Mean	**S. D.**	**Mean**	**S. D.**
Language Arts Core Battery (N = 6)	**45.300**	**3.078**	**46.967**	**2.783**
Mathematics Core Battery (N = 6)	**42.017**	**3.499**	**41.567**	**3.596**
Reading Core Battery (N = 6)	**39.750**	**3.135**	**39.833**	**2.159**
Total Core Battery (N = 18)	**42.356**	**3.844**	**43.117**	**3.718**

Questionnaire Results

Themes in Qualitative Data

Data analysis of participant responses on post training questionnaires yielded the following themes concerning performance assessment.

Theme 1: Confidence in Test Taking. Several teachers reported that strategies implemented in their classrooms following the staff development sessions raised the confidence level of their students and allowed them to share their ideas in a non-threatening way. Reports

from staff development session leaders as well as the final evaluation forms show that some teachers reported that students said they felt prepared for the test. Students said the test was like things they had done in the classroom and they knew how to do it. One teacher reported, "I saw students using the checklist. I saw students revisiting the text. They really understood what they were doing. They were not as nervous as I thought they would be. They were fully prepared." Comments from teachers indicate that students were more confident during the testing because they were prepared for and familiar with the test after participating in strategies from the performance assessment staff development sessions.

Theme 2: Strategies Used During the Test. Reports from staff development session leaders as well as the final evaluation forms show that teachers noticed students using some of the strategies implemented in the classroom from the staff development sessions during the administration of the test. Teachers reported that students wrote more during the performance portion of the test than they had during previous years. Teachers also reported that students returned to the text to mark details to use in their written responses to the reading. Teachers also said that students appeared to use the checklist to ensure that they had included details from the reading in their written responses. One teacher said, "During the performance portion of the ITBS/TAP, I noticed my students underlining words/phrases, visiting the checklist before writing, and revisiting the text and checklist as they wrote." Another teacher reported, "I noticed the students revisiting the text, underlining, and checking over their work using the checklist during the test." It appears from the teachers' comments on the final evaluations that students utilized strategies for completing performance tasks while taking the performance portion of the ITBS/TAP.

Theme 3: Changes in Teachers' Practices. On the final evaluation questionnaire teachers noted that they had implemented many of the strategies that they had seen modeled in the staff development sessions. Two of the most frequently implemented strategies were focused sharing and small group activities. Several of the teachers reported that these activities raised the confidence level of their students and allowed them to share their ideas in a non- threatening way. One teacher reported, "In small group activities, the students were able to interact more confidently and openly in sharing ideas and thoughts."

Other teachers initiated even more in depth implementation from the sessions. Many planned additional performance units for reading and math. Others reported that they developed checklists with their students as well as developing and using rubrics. One teacher said, " . . . developing performance units in my classroom has allowed me to integrate math, social studies, science and art into reading/language arts . . . " Teachers reported that they had initiated several changes in their classroom practices based on strategies they had participated in during the staff development sessions.

Theme 4: Developing Higher Order Thinking Skills. Teachers reported that implementing performance units in their classrooms helped students develop higher order thinking skills. Teachers implemented strategies, which engaged students in critical thinking, writing, stating opinions, and self-assessment. One teacher said, ". . . developing performance units in my classroom has helped improve students' thinking and writing skills . . . " Teachers also reported that small group activities and collaborative learning situations enabled the students to interact more confidently and openly in sharing ideas and thoughts. Teachers' comments reflect that these practices helped their students develop and share opinions, gain insights from each other, and improve their thinking skills.

Discussion

Based on the results of the present study, a number of tentative conclusions can be made. First, when mathematics performance assessment scores and language performance scores are analyzed separately, the proposed hypothesis that performance task instruction will increase performance assessment scores has not been supported. However, when total performance assessment scores are analyzed, there is a significant difference or improvement between scores for 1998 and 1999. A larger sample size would possibly yield significant differences between the uncombined mathematics and language performance scores. Second, this data set shows the absence of a significant relationship between performance assessment scores and core battery scores. This lack of relationship implies that-- contrary to the findings of Elliot and Fuchs(1997)-- core battery scores are unaffected by students performance on the performance assessment part of the ITBS/TAP and vice versa. Even though a statistically significant correlation was not found between performance assessment scores and core battery scores, from 1997 to 1999, the districts' score sets have shown a gradual increase in the correlation coefficient each year. The greatest correlation between performance assessment and core battery scores occurred in 1999. Variance in the performance assessment scores accounted for approximately 14% of the variance in the core battery scores.

Although analyses indicate that performance task instruction did not indicate a significant increase in scores, inspection of the means shows an increase (see Table 1). This increase in scores could be due, in part, to the performance task instruction in which the teachers received training. Teacher perceptions about the relationship student performance and Integrated Assessment training has been documented in this data set.

Improvement in both the multiple choice portion and the performance assessment portion of the ITBS occurred after the integrated assessment treatment was administered, though no statistical significance was found in the improvement or the correlation of score sets. General inspection of the means supports that score on both multiple choice and performance test items improved after students had been introduced to ideas of performance based instruction. When examining this improvement in mean scores in conjunction with the teachers' responses to the post training questionnaire, clearly teachers felt using performance based instruction a worthy endeavor for students academically. Since students began to respond to multi-stepped instructions, regularly reacted to literature, and practiced returning to text during daily lessons, the tendency to practice these habits during the testing situation affected scores. Teachers felt the increased presence of writing in the students' daily routine had some positive affect on the students' ability to construct responses or write during the test. Still, this increase in students' scores shows only one positive impact of performance based instruction.

More importantly is the theme of developing higher order thinking skills which emerged. Teachers felt that using performance task instruction encouraged critical thinking. Teachers saw students write more, state opinions, and self assess more. Instead of regurgitating facts with only a knowledge or comprehension level of understanding, teachers noticed that students were participating in collaborative learning situations, gaining insight from each other, and interacting more confidently. As educators examine what is important for students to know and do, surely collaboration, thinking critically, and expressing ideas, is superior to acts of fact recall.

Teachers also seemed to feel that performance based instruction was a worthy endeavor for themselves professionally. Teachers professed to engage more readily in writing original curriculum pieces after attending the Integrated Assessment Module. These original teaching

materials included cross-curricular content similar to the materials teachers receive during Integrated Assessment. Teachers stated they used cooperative groups more often as a result of their training. Rubrics and checklists became a regular part of some teachers' instruction. The module helped teachers see how units of instruction could incorporate math social studies, science, art, and language into literature based teaching while continuing to address the state standardized assessment demands.

One additional benefit of teaching with performance based instruction deals with test anxiety. Teachers professed that the students' experience of test anxiety was reduced after being taught with techniques teachers had learned in the Integrated Assessment training module. Likewise, teachers stated that students experienced an overall feeling of success during the testing situation.

Conclusion

The PREPS Integrated Assessment Module of training seemed to positively effect teaching and learning in this Mississippi school district. Not only did students' scores improve after the training, but teachers' professed their learning through behavior changes like the writing of original curriculum and the integration of content areas into larger, more thematic units. Both teachers and students grew from using performance-based instruction and assessment. The district has elected to continue investing time, energy, and dollars in performance based professional development for teachers. This district seems committed to steering accountability measures away from multiple choice tests, helping students construct knowledge and understanding in the information age, and preparing students for a highly productive, globally competitive world through its use of performance task instruction.

References

Borko, H., Mayfield, V., Marion, S., Flexer, R., & Cumbo, K. (1997). Teachers' developing ideas and practices about mathematics performance assessment: Successes, stumbling blocks, and implications for professional development. Teaching & Teacher Education, 13, 259-278.

Brooks, J. G., & Brooks, M. G. (1993). In search of understanding: The case for constructivist classrooms. Alexandria, VA: Association for Supervision and Curriculum Development.

Elliot, S. N., & Fuchs, L. S. (1997). The utility of curriculum-based measurement and performance assessment as alternatives to traditional intelligence and achievement tests. School Psychology Review, 26, 224-233.

Gardner, H. (1994). Multiple intelligences: A view from the arts. Issues 1994, 5-22. Based on a talk delivered to a conference of the Art Educators of New Jersey, October 1993.

Givens, K. (1997). Performance assessment tests: A problematic panacea. Contemporary Education, 69, 27-29.

Haertel, E. H. (1999). Performance assessment and education reform. Phi Delta Kappan, 80, 662-666.

Hakel, M. D. (1998). Beyond multiple choice: Evaluating alternatives to traditional testing for selection. Mahwah, NJ: Lawrence Erlbaum Associates.

Hinckley, J. (1998). What gets tested gets taught. Teaching Music, 6, 6-7.

Khattri, N., Reeve, A. L., & Kane, M. B. (1998). Principles and practices of performance assessment. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Madaus, G. F., & O□Dwyer, L. M. (1999). A short history of performance assessment. Phi Delta Kappan, 80, 688-695.

Martin, M., & Miller, G. (1996). Instructional improvement through performance assessment. Thrust for Educational Leadership, 25, 10-12.

Miller, M. G. (1999). Writing abstracts on a central theme. Journal of College Science Teaching, 28, 397-400.

Mississippi Department of Education. (1995). Mississippi report card 1995. Jackson, MS: Author.

Mississippi Department of Education, Office of Educational Accountability. (1997). Mississippi report card 1997. Jackson, MS: Author.

Mississippi Department of Education. (1998). Mississippi report card 1998. Jackson, MS: Author.

Mississippi Department of Education. (1999). Mississippi report card 1999. Jackson, MS: Author.

Muscella, D., Paget, K., Gibbel, E. (1994). A conversation between teachers and researchers. Hands on!, 17, 1-6.

Appendix

Summary of Teacher Responses from the Post-treatment Questionnaire

Theme 1: Confidence in Test Taking	Theme 2: Strategies Used During Testing	Theme 3: Changes in Teachers' Practices	Theme 4: Developing Higher Order Thinking Skills
They really understood what they were doing They were not as nervous as I thought they would be. They were fully prepared.	I noticed the students revisiting the text, underlining, and checking over their work using the checklist during the test.	Getting ideas to help students achieve better on ITBS integrated assessment.	Some students went back to the text and underlined ideas to help them when writing.
The focused sharing really helped some of my shy students. To get in a small group and voice their opinions without being judged or thinking you might be wrong, gave my students more confidence.	I saw students using the checklist. I saw students revisiting the text.	The PREPS units gave me a guide in developing my own units.	The students really liked doing the unit. It helped them follow directions and enabled them to go back to make sure that all assignments were completed.
In small group activities, the students were able to interact more confidently and openly in sharing ideas and thoughts.	Students revisited the text and underlined information during the test.	Mini-lessons in context: the mini-lesson was very helpful with specific skills needing addressing according to the benchmark.	This thoroughly prepared students for written assessment.
Students said they felt prepared for the test.	Students underlined and color coded.	I developed PREPS units to go along with geometric shapes.	This really allowed the students to self-assess and see how they are assessed.
These activities raised the confidence level of the students and allowed them to share their ideas in a non-threatening way.	During the performance portion of the ITBS test, I noticed my students underlining, words/phrases, visiting the checklist before writing, revisiting the text and checklist as they wrote.	I developed preps units to go along with reading units.	Having students engage in focused sharing served as reinforcement of thought process before writing.
Students said the test was like things they had done in the classroom and they knew how to do it.	Some used checklist to help with performance.	I used checklist and rubrics by using performance assessments.	The students enjoyed working together on the bear unit. They were able to build on one another's ideas.
	Students used colored pencils for multi-step questions. Students enjoyed this.	Small groups are successful.	Slower learners gained insights from the accelerated learners.
		The students worked cooperatively in small groups, which allowed me to assist slower learners.	The impact of developing performance units in my classroom has: (a) helped, improve students thinking and writing skills. . . (c) helped students become more focused on assignments.
		The impact of developing performance units in my classroom has: . . . (b) allowed me to integrate math, social studies, science and art into reading/language arts.	Students wrote more than they had during previous years.

BEST COPY AVAILABLE